Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

REMARKS/ARGUMENTS

These remarks are made in response to the Final Office Action of February 7, 2005 (Office Action). As this response is timely filed within the 3-month shortened statutory period, no fee is believed due.

The Examiner has rejected Claims 1, 3-4, 15, 17-18, 29-30, 32 and 33 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,182,039 to Rigazio, et al. (Rigazio), in view of U.S. Patent No. 5,133,012 to Nitta (Nitta). Claims 2, 16, and 31 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view Nitta and further in view of U.S. Patent No. 6,208,966 to Bulfer (Bulfer). Claims 6 and 20 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view Nitta and Bulfer and further in view of U.S. Patent No. 5,829,000 to Huang, et al. (Huang).

Claims 10, 14, 24, and 28 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view Nitta, Bulfer, and Huang, and further in view of U.S. Patent No. 4,696,042 to Goudie (Goudie). Claims 8, 12, 22, and 26 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view of Nitta, Bulfer and Goudie. Claims 5, 19, and 34 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view of Nitta and further in view of Huang. Claims 9, 13, 23, and 27 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view of Nitta and Huang and further in view of Goudie. Claims 7, 11, 21, 25, 35, and 36 have been rejected under 35 U.S.C. § 103(a) as being unpatentable in view of Rigazio in view of Nitta and further in view of Goudie.

Independent Claims 1, 15, 29, and 30 have been amended to more clearly delineate Applicant's invention. The amendments are fully supported in the specification, and no new matter has been added as a result of the amendments.

13

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

I.   Applicant's Invention

Applicant's invention is directed to a system and method for improving speech recognition accuracy, especially with respect to the recognition of characters such as numbers and words that are characterized by disparate readings or renderings. (Specification, p. 6, lines 2-5.) The numerical symbol "0" provides a familiar example of character having a different reading or rendering. (Specification, p. 19, lines 6-26.) Some individuals typically pronounce the symbol by saying "zero." Others, however, commonly render the symbol as "oh." Still others verbally render the symbol as "aught." Note, in particular, that none of these different readings or renderings of the same character have any acoustic or phonetic similarity. Applicant's invention is particularly well suited for recognizing characters that lend themselves to disparate acoustic renderings.

Note also that different renderings of the same character or word are virtually inevitable when two speakers speak a different language, and each renders the same character in his or her own language. (Specification, p. 14, lines 7-13.) Moreover, not only do readings or renderings varying among speakers, but, so too, renders of a word or character can vary at different times with the same speaker. (See Specification, p. 5, lines 5-7.)

One embodiment of Applicant's invention pertains to a speech recognition system that includes correspondence information in which is stored a correspondence between recognized words and a plurality of speech element arrays, each array comprising associated rendering information for expressing pronunciation of the recognized words. The associated rendering information, more particularly, comprises at least one set of alternate renderings of a recognized word. (Specification, p. 14, lines 7-23.) The speech recognition system recognizes a

14

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

recognizable word from a received user-spoken utterance by comparing a speech element array generated from the user-spoken utterance with the plurality of speech element arrays in said correspondence information.

In a dialog of a single person occurring within a certain period of time, the generated speech element array corresponds to one of the plurality of speech element arrays. A pronunciation prediction probability corresponding to one of the plurality of speech element arrays is lowered by uniquely associating with the person one alternate rendering from the set of alternate renderings. (Specification, p. 15, line 13 – p. 16., line 5; p. 16, lines 16-25.) For example, if during one phase of a speech recognition of the characters "740," it is determined that the speaker has rendered the third digit as "zero," the other possible renderings (e.g., "oh" or "aught") are not further considered as possible renderings during the remainder of the session. (Specification, p. 15, lines 13-22.)

A similar result obtains for more complex character combinations such as 3xx. If during a session, a user articulates the last two characters by uttering "x" twice, the alternate rendering "double x" is excluded from consideration during the remainder of the session.

Another embodiment of Applicant's invention is directed to a method of speech recognition for use within a dialog of a single person occurring in a certain period of time. The method includes receiving a first user-spoken utterance and generating a first speech element array from the first user-spoken utterance. The method further includes searching correspondence information, the correspondence information associating recognizable words with a plurality of speech element arrays that each comprise associated rendering information for expressing pronunciation of the recognized words. The associated rendering information, moreover, comprises at least one set of alternate renderings of a recognized word.

15

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

Additionally, the method includes generating a first recognized word by comparing the first speech element array and the plurality of speech element arrays in the correspondence information, and lowering a pronunciation prediction probability of one of the plurality of speech element arrays that differs from the first speech element array. The latter is achieved by uniquely associating with the person one alternate rendering from the set of alternate renderings.

Other alternate renderings are excluded from further consideration during the session. The method also includes receiving a second user-spoken utterance and generating a second speech element array from the second user-spoken utterance, searching the correspondence information wherein the other alternate renderings are excluded from consideration, and generating a second recognized word by comparing the second speech element array and the plurality of speech element arrays in the correspondence information.

II. The Combination of Rigazio And Nitta Fail To Render The Claims Obvious

As noted above, Claims 1, 3-4, 15, 17-18, 29-30, 32 and 33 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Rigazio in view of Nitta. A claim can be deemed *prima facie* obvious only if each limitation recited in the claim is taught or suggested by the prior art. *In re Royka*, 490 F.2d 981 (CCPA 1974). Applicants respectfully maintain that the cited references, even when combined, fail to teach or suggest each limitation recited in the claims.

A. Rigazio fails to address non-acoustic attributes in speech recognition and does not suggest information comprising at least one set of alternate renderings of a recognized word

Rigazio is directed to a speech recognizer incorporating a language model that "reduces the number of acoustic pattern matching sequences that must be performed

16

by the recognizer." (Abstract.) (emphasis supplied.) By reducing the number of matching sequences, the language model in Rigazio is intended to speed up a recognition event and save memory space. (Col. 3, lines 57-64; Col. 3, line 66 – Col. 4, line 4.) To reduce the number of matches, the language model incorporates a set of confusable classes, which are defined as "sets of letters having sounds that the recognizer conventionally has difficulty discriminating among." (Col. 6, lines 2-4.)

Rigazio is intended to operate in the context of recognizing names, where taking account of syntax is not helpful for resolving "acoustic confusability." (Col. 4, line 55 – Col. 5, line 22.) Accordingly, Rigazio "pre-determines" the different sets of letters that are likely to be acoustically confusing. This leads to a language model that, as described in Rigazio, can be represented using several different data structures (e.g., N-gram and tree). (Col. 6, lines 5-15.) The reduced matching and memory requirements are achieved, according to Rigazio, because each of the various data structures is smaller than the dictionary-matching data structure of other recognizers. (Col. 6, line 55 – Col. 7, line 17.)

What is striking about Rigazio in the present context, is that Rigazio is exclusively focused on resolving "acoustic confusability;" that is, discriminating between different utterances that sound similar. Rigazio specifically states that the "present invention improves the recognizer's ability to discriminate between words within the language that are very similar in sound. (Col. 3, lines 57-59.) Rigazio, in this sense, is the precise opposite of Applicants' invention.

In focusing on different words that sound similar, Rigazio does not even consider that the same character or word may have multiple, albeit entirely disparate, renderings. In exclusively addressing acoustic and phonetic attributes of similar sounding words from the same language, Rigazio provides no insight into non-phonetic disparities among the same characters and words. Accordingly, Rigazio does not teach or suggest a speech recognition system that includes correspondence

17

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

information for providing a correspondence between recognized words and a plurality of speech element arrays that each comprise associated rendering information, the associated rendering information comprising at least one set of alternate renderings of a recognized word as recited in independent Claim 1 as amended.

Rigazio similarly fails to teach or suggest other features of Applicant's invention. Rigazio does not teach or suggest a speech recognition system wherein, during one person's dialog within a certain period of time, a speech element array is generated corresponding to one of a plurality of speech element arrays while a pronunciation prediction probability corresponding to one of the plurality of speech element arrays is reduced by uniquely associating with the person an alternate rendering from the set of alternate renderings. These features are also recited in independent Claim 1, as amended.

Similarly, Rigazio does not teach or suggest a method that includes searching correspondence information that associates recognizable words with a plurality of speech element arrays that each comprise associated rendering information for expressing pronunciation of the recognized words, the associated rendering information comprising at least one set of alternate renderings of a recognized word as recited in each of independent Claims 15, 29, and 30, as amended. Nor does Rigazio teach or suggest a method whereby a pronunciation prediction probability of one of the plurality of speech element arrays is altered by uniquely associating with a user one alternate rendering from the set of alternate renderings and excluding other alternate renderings from further consideration during a given dialog or period of time, as also recited in amended independent Claims 15, 29, and 30.

Rigazio is incapable of accomplishing what Applicant's invention achieves. For example, Rigazio's pre-determined sets of acoustically confusable letters can not be adapted to provide information concerning alternative renderings for the same

18

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

word or character. It follows that Rigazio's language model can not be modified to associate a particular one of two or more alternative renderings with a person during a dialog so as to affect a prediction probability as recited in amended independent Claim 1.

B. Nitta does not suggest altering a pronunciation prediction probability by uniquely associating with a user an alternate rendering from a set of alternate renderings

The Examiner correctly points out in paragraph one of the Office Action that Rigazio does not teach or suggest altering a pronunciation prediction probability. It is contended, however, that this feature is found in Nitta.

Nitta is directed to a speech recognition system that selects a word from among a plurality of candidate words based on their respective scores. The selected word is the output of the system and represents a recognized word. (Col. 2, lines 57 – 62.)

The scores are based upon a Bayesian statistical measure of the similarity of received speech and a stored phonetic features (a "long-term strategic score") and a consideration of the segment environment of phonetic segment of the received speech (a "short-term strategic score"). Col. 7, line 11 – Col. 8, line 66.) The score determines the degree of closeness of a received speech segment and a possible word match.

The score-based matching in Nitta is thus based on the phonetic features of a received speech utterance. This is entirely distinct, however, from the prediction probability adjustments recited in each of independent Claims 1, 15, 29, and 30, as amended. As described above, the prediction probability is adjusted in Applicant's invention in response to the unique association of one of alternate renderings of a word or character with a particular user during a given dialog or session. Other alternatives are excluded, thus a subsequent speech recognition event has a smaller

19

Appln. No. 09/846,122   IBM Docket No. JP9-2000-0040US1
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

set of possible matches. This is not based, however, on a statistical or other type of measurement of the phonetic characteristics of the character or word, but rather by the particular user being associated with a particular one of the alternate renderings for the same character or word.

An example, as described by Applicant, is where a user is asked to render a number (e.g., a customer number) at a point in the dialog. The customer may render the number 740 by saying "7" then "4" and then "zero." From that point on, the alternate renderings of the last digit, "oh" and "aught," are excluded from consideration according to Applicant's system and method. During subsequent portions of the same dialog, therefore, the verbal rendering of the character "0" entails matching the character with less than all the possible renderings. The resulting change in the underling prediction probabilities, accordingly, is not derived from a measure of phonetic similarities, as in Nitta. Indeed, the change has no relation to phonetic attributes since "zero," "oh," and "aught" have no phonetic similarity at all. Instead, the distinction is the user's opting to use one particular rendering over the alternate possible renderings.

In so far as Nitta is based on a statistical measure of phonetic similarities, Nitta neither teaches nor suggests any type of measure or alteration of prediction probabilities comparable to that recited in independent Claims 1, 15, 29, and 30. It follows, therefore, that neither Nitta nor Rigazio teach or suggest this feature of Applicants invention.

Rigazio and Nitta thus fail to teach or suggest each feature of independent Claims 1, 15, 29, and 30, as amended. Applicant respectfully submits, therefore, that the cited references fail to render the independent claims *prima facie* obvious. Applicant further respectfully submits that, whereas each dependent claim adds additional features, each is thus likewise not rendered *prima facie* obvious by the cited references.

20

Appln. No. 09/846,122
Amendment dated Apr. 07, 2005
Reply to Office Action of Feb. 07, 2005
Docket No. 6169-237

IBM Docket No. JP9-2000-0040US1

CONCLUSION

Applicant respectfully requests that the rejection of Claims 1-36 be withdrawn for the reasons stated herein. Applicant believes that this application is now in full condition for allowance, which action is respectfully requested. Applicant requests that the Examiner call the undersigned if clarification is needed on any matter within this Amendment, or if the Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Date: April 7, 2005
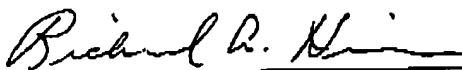
Gregory A. Nelson, Registration No. 30,577
Richard A. Hinson, Registration No. 47,652
Brian K. Buchheit, Registration No. 52,667
AKERMAN SENTERFITT
Customer No. 40987
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000

21